

Вдосконалення моделі розпізнавання об'єктів на зображеннях з використанням згорткової нейронної мережі

Б. П. Книш, Я. А. Кулик

Розглянуто модель розпізнавання об'єктів на зображеннях за допомогою згорткових нейронних мереж та дослідження ефективності процесу на основі моделей із навчанням глибоких шарів згорткових нейромереж. Існують об'єктивні труднощі, пов'язані з визначенням оптимальних характеристик нейронних мереж, тому є проблема перенавчання нейромережі. Усунення перенавчання шляхом визначення лише оптимальної кількості епох недостатнє, оскільки не забезпечує високу точність.

Визначено вимоги до набору зображень для навчання та перевірки моделі. Даним вимогам найбільш відповідає набір зображень INRIA (Франція).

Встановлено, що GoogLeNet (США) є навченою моделлю і може виконувати розпізнавання об'єктів на зображеннях, проте надійність розпізнавання об'єктів недостатня. Тому виникає необхідність підвищення ефективності розпізнавання об'єктів на зображеннях. Доцільно використати архітектуру GoogLeNet для створення спеціалізованої моделі, яка за рахунок зміни параметрів та перенавчання деяких шарів дозволить краще проводити процес розпізнавання об'єктів на зображеннях.

Виконано навчання десяти моделей з використанням таких параметрів: швидкість навчання, число епох, алгоритм оптимізації, вид зміни швидкості навчання, коефіцієнт гатта чи power, попередньо навчена модель.

Розроблено згорткову нейронну мережу для підвищення точності та ефективності розпізнавання об'єктів на зображеннях. Визначено оптимальні параметри навчання нейромережі: швидкість навчання – 0,000025, число епох – 100, коефіцієнт power – 0,25 тощо. Отримано підвищення точності на 3 %, яке дозволяє стверджувати про правильність вибору архітектури розробленої мережі та підбору параметрів. Це дає можливість використовувати дану мережу для практичних задач розпізнавання об'єктів на зображеннях.

Ключові слова: обробка зображень, розпізнавання об'єктів, згорткові нейронні мережі, безпілотний літальний апарат.

1. Вступ

Обробка зображень має надзвичайно важливе значення в сучасній науці та практиці, тому постійно розвивається та вдосконалюється. Обробку зображень можна використовувати в багатьох галузях виробництва, а саме в точному землеробстві (агромоніторингу), системах безпеки, контролі якості тощо. У перелічених галузях задіяні системи технічного зору, робототехнічні комплекси, безпілотні літальні апарати (БПЛА), системи відеоспостереження, веб-сервіси та мобільні застосування для ідентифікації та пошуку.

Одним з видів обробки зображення є розпізнавання об'єктів на зображеннях, яке широко застосовується в промисловості, мистецтві, медицині, космосі, при керуванні процесами, автоматизації та в багатьох інших галузях [1]. Розпізнавання об'єктів на зображеннях передбачає віднесення вихідних даних до певного класу шляхом виділення суттєвих ознак. Ці ознаки характеризують вихідні дані з загального масиву несуттєвої інформації.

Існує велика кількість методів розпізнавання об'єктів на зображеннях, серед яких найбільш поширеними є методи Random Forests, boosting-методи, а також нейромережеві методи, зокрема згорткові [2–6].

До методів розпізнавання об'єктів висуваються певні вимоги, а саме:

- відповідність розпізнаного об'єкта реальному;
- висока швидкодія;
- стійкість до помилок;
- висока точність.

Тому виникає необхідність аналізу методів розпізнавання об'єктів на зображеннях та вибору оптимального згідно вищенаведених вимог, зокрема високої точності. Також варто врахувати параметри, які характеризують ці методи, зміна яких здійснює безпосередній вплив на точність, швидкодію та загальну ефективність процесу розпізнавання об'єктів.

Сучасним актуальним напрямком виробництва є розвиток точного землеробства, яке ґрунтується на результатах агромоніторингу. Ці дані, отримані з відеокамер БПЛА, дозволяють здійснювати оцінку зібраного врожаю, контролювати маршрути переміщення сільськогосподарської техніки, прогнозувати врожайність тощо. Важливим критерієм при цьому виступає здатність БПЛА уникати зіткнень із близькими предметами, визначати положення в просторі, напрям та траєкторію польоту за рахунок надходження вхідних даних про розпізнані об'єкти.

Ефективність цих систем визначається точністю розпізнавання об'єктів, для оцінки якої необхідні експериментальні дослідження.

2. Аналіз літературних даних та постановка проблеми

В роботі [2] розглядається метод Random Forests для розпізнавання багатьох класів об'єктів, який характеризується високою точністю, стійкістю до перенавчання та легко прискорюється при використанні паралельних обчислень. Але залишаються невирішені питання, пов'язані з відсутністю візуальної інтерпретації процесу і складністю пояснень їх рішень, а також високою чутливістю до шумів на зображеннях. Це викликає труднощі, пов'язані з високими вимогами до відсутності шумів на зображеннях та неможливістю отримання пояснення результату. В роботах [3, 4] наводяться результати розпізнавання об'єктів на зображеннях за допомогою boosting-методів, зокрема Adaboost. Показано високу швидкість та ефективність роботи, а також адаптивність до конкретного застосування. Проте є складнощі, пов'язані з перенавчанням при наявності шумів у вхідних даних, великою кількістю ознак зображення, а також з необхідністю значної кількості даних для навчальної вибірки. Це робить дослідження затратними та обмежує використання даних методів при роботі з зображеннями низької якості. В роботі [5] наводяться

результати розпізнавання об'єктів на зображеннях за допомогою нейромережових методів. Показано здатність до навчання системи для виділення ключових характеристик об'єктів з навчальних вибірок. Проте дані методи вимагають використання ансамблю нейронних мереж, допоміжних методів виділення сюжетної частини зображення, а також їх архітектури є вкрай чутливими до зовнішніх впливів. Причиною цього можуть бути труднощі, пов'язані з обчислювальною складністю та якістю попередньої обробки навчальних і робочих даних. Це робить використання даних методів для певних задач не ефективними. В роботі [6] наводяться результати реальних задач розпізнавання об'єктів на зображеннях за допомогою нейромережових методів. Показано, що вхідні дані можуть бути представлені в будь-якому порядку, що не впливає на мету навчання. Проте дані методи вимагають врахування великої кількості параметрів, оскільки зображення в реальних задачах розпізнавання мають велику розмірність. Причиною цього можуть бути труднощі, які спричинені вимогами більшої тренувальної вибірки. Це збільшує час і обчислювальну складність процесу навчання, що обмежує використання даного методу.

Варіантом подолання вищенаведених труднощів, пов'язаних з недостатньою точністю, ефективністю та швидкістю, може бути використання методів розпізнавання об'єктів на зображеннях на основі згорткових нейронних мереж [7, 8]. Саме такий підхід використаний у роботі [9], де використовуються багатоспектральні дані, отримані з супутника, тоді як відеокамери БПЛА забезпечують мультиспектральні дані. Також схожий принцип реалізований в роботі [10], де аналізуються параметри навчання та надаються рекомендації по зміні архітектури нейромережі, проте ці рекомендації містять загальний характер без аналізу конкретних застосувань, зокрема для задач розпізнавання.

Все це дає підстави стверджувати, що доцільним є проведення дослідження по підвищенню ефективності навчання нейронної мережі, що дозволить значно покращити точність розпізнавання об'єктів на зображеннях.

3. Мета та задачі дослідження

Метою роботи є вдосконалення моделі згорткової нейронної мережі для розпізнавання об'єктів на зображеннях та вибору параметрів навчання цієї мережі. Це дасть можливість отримати нову нейромережу із підвищеною точністю для розпізнавання об'єктів на зображеннях, яка може бути використана як попередньо навчена нейромережа для інших задач.

Для досягнення мети були поставлені наступні завдання:

- дослідити моделі нейромереж на основі набору INRIA;
- оцінити модель Inria-9.

4. Матеріали та методи дослідження

Дослідження розпізнавання об'єктів на зображеннях проводилось за допомогою відповідних методів на основі згорткових нейронних мереж з врахуванням параметрів навчання нейромережі. Для перевірки ефективності цих методів використовувався набір INRIA, який містить велику кількість зображень із розміченими групами пікселів та визначеними класами об'єктів. INRIA містить

зображення, які отримані з відеокамер, прикріплених на БПЛА при зйомці з висоти декількох сотень метрів [11]. Дослідження проводилось за допомогою програмного середовища DIGITS з використанням середовища Caffe, яке призначено для глибокого навчання нейромережі з врахуванням швидкодії та модульності при розробці моделі. Поєднання цих середовищ дозволяє швидко навчати нейронні мережі з глибокими шарами та використовується для задач класифікації, сегментації зображень [12] та розпізнавання об'єктів на них. DIGITS містить попередньо навчену модель GoogLeNet, яка характеризується адаптованими параметрами для розпізнавання об'єктів на зображеннях (табл. 1, 2) та має гнучку архітектуру (рис. 1).

Архітектура GoogLeNet складається з 22 шарів (27 шарів, якщо враховувати шари об'єднання) і частина цих шарів складається з 9 початкових модулів. Причому їх параметри можуть змінюватись в процесі навчання. На вхід подається зображення з палітрою RGB розміром 224×224 . Розмір фільтра першого шару становить 7×7 . Використовується ядро розміром $1 \times 1 \times 256$. Функція активації на виході – Softmax, а в шарах – ReLU, яка дозволяє в 6 разів збільшити швидкодію. В порівнянні з аналогічними моделями [13, 14], GoogLeNet містить в 12 разів менше параметрів, глибина мережі збільшена до 22 шарів без додаткового залучення обчислювальних ресурсів [15].

Таким чином, архітектуру GoogLeNet було використано як базову для створення спеціалізованої моделі FCN-GoogLeNet шляхом додавання повнозв'язного згорткового шару через внесення таких змін в DIGITS:

- додані шар даних, який отримує навчальні зображення та мітки, та шар перетворення, який застосовує збільшення даних в реальному часі;
- доданий шар нормалізації даних;
- додана повнозв'язна згорткова мережа (FCN), яка здійснює вилучення характеристик та прогнозування класів об'єктів та меж полів на квадрат сітки;
- доданий шар похибки, який одночасно вимірює два значення прогнозування;
- після визначення розміру вхідного зображення встановлюється випадкове число, яке визначає на скільки потрібно зменшити вхідне зображення;
- додані параметри для доповнення даних, що визначають якою мірою випадкові перетворення (зміщення пікселів, перевертання зображення тощо) повинні застосовуватися до вхідних зображень;
- доданий шар, який використовує лінійну комбінацію двох окремих функцій втрат для обрахунку загальної функції втрат для оптимізації;
- видалені шари вхідних і вихідних даних та pooling шар [16].

Вибір алгоритму оптимізації моделі FCN-GoogLeNet визначається особливостями розпізнавання об'єктів на зображеннях, для якої необхідно мати хорошу збіжність алгоритму, а для практичного використання – високу швидкодію.

Порівняння алгоритмів [17] показує, що для задачі розпізнавання об'єктів на зображеннях найкращі результати швидкодії показує Adam (збільшення на 10–50 %). Також цей алгоритм демонструє хорошу збіжність.

Таблиця 1
Параметри моделі GoogLeNet

Тип	Розмір па-тчу/крок	Розмір ви-воду	Гл иб ина	#1 ×1	#3×3 до згор-тки	#3 ×3	#5×5 до згор-тки	#5 ×5	Філь-три 1×1	Ваги	Мате-матич-ні опе-рації
convolu- tion	7×7/2	112×112×64	1	—	—	—	—	—	—	2,7К	34М
max pool	3×3/2	56×56×64	0	—	—	—	—	—	—	—	—
convolu- tion	3×3/1	56×56×192	2	—	64	19 2	—	—	—	112К	360М
max pool	3×3/2	28×28×192	0	—	—	—	—	—	—	—	—
incep- tion (3a)	—	28×28×256	2	64	96	12 8	16	32	32	159К	128М
incep- tion (3b)	—	28×28×480	2	128	128	19 2	32	96	64	380К	304М
max pool	3×3/2	14×14×480	0	—	—	—	—	—	—	—	—
incep- tion (4a)	—	14×14×512	2	192	96	20 8	16	48	64	364К	73М
incep- tion (4b)	—	14×14×512	2	160	112	22 4	24	64	64	437К	88М
incep- tion (4c)	—	14×14×512	2	128	128	25 6	24	64	64	463К	100М
incep- tion (4d)	—	14×14×528	2	112	144	28 8	32	64	64	580К	119М
incep- tion (4e)	—	14×14×832	2	256	160	32 0	32	128	128	840К	170М
max pool	3×3/2	7×7×832	0	—	—	—	—	—	—	—	—
incep- tion (5a)	—	7×7×832	2	256	160	32 0	32	128	128	1072К	54М
incep- tion (5b)	—	7×7×1024	2	384	192	38 4	48	128	128	1388К	71М
avg pool	7×7/1	1×1×1024	0	—	—	—	—	—	—	—	—
dropout (40%)	—	1×1×1024	0	—	—	—	—	—	—	—	—
linear	—	1×1×1000	1	—	—	—	—	—	—	1000К	1М
softmax	—	1×1×1000	0	—	—	—	—	—	—	—	—

Основними показниками ефективності навчання нейромережі, що визначалися в процесі дослідження, було обрано такі характеристики [18]:

– влучність (precision) – це відношення правильно розпізнаних об’єктів до загальної кількості передбачуваних чи істинних об’єктів:

$$Precision_{val} = \sum_{k=1}^{N_{val}} \frac{N_{TP_k}}{N_{TP_k} + N_{FP_k}}, \quad (1)$$

де N_{TP} визначається як кількість правильно розпізнаних об'єктів на зображенні; N_{FP} визначається як кількість помилково розпізнаних об'єктів; N_{val} – кількість зображень у перевіірочній вибірці; k – поточне зображення;

– чутливість (recall) – це відношення правильно розпізнаних об'єктів до загальної кількості об'єктів на зображеннях:

$$Recall_{val} = \sum_{k=1}^{N_{val}} \frac{N_{TP_k}}{N_{TP_k} + N_{FN_k}}, \quad (2)$$

де N_{FN} визначається як кількість помилково нерозпізнаних об'єктів;

– оцінка усередненої точності (mean Average Precision) – це спрощена оцінка математичного сподівання на основі добутку влучності та чутливості, яка показує наскільки мережа чутлива до потрібних об'єктів і стійка до помилок:

$$mAP = Precision_{val} \times Recall_{val}. \quad (3)$$

Таблиця 2

Параметри шарів моделі GoogLeNet

Тип	Кількість нейронів	Загальна кількість зв'язків нейронів у шарі	Кількість зв'язків з наступним шаром
convolution	20M	402M	3K
max pool	1M	125M	3K
convolution	1,2M	944M	784
max pool	401K	78M	784
inception (3a)	200K	78M	784
inception (3b)	376K	295M	196
max pool	200K	18M	196
inception (4a)	100K	9M	196
inception (4b)	100K	19M	196
inception (4c)	100K	19M	196
inception (4d)	103K	20M	196
inception (4e)	163K	31M	49
max pool	81K	998K	49
inception (5a)	40K	1,9M	1
inception (5b)	1024	1024	1
avg pool	25K	40	1
dropout (40%)	1024	1024	1
linear	1000	1000	1
softmax	1000	1000	—

Для оцінки ефективності навчання нейромережі визначаються оптимальні параметри нейронної мережі. Цими параметрами є тривалість навчання (число епох), алгоритм оптимізації (адаптивна миттєва оцінка (Adam)), вид зміни шви-

дкості навчання, коефіцієнт gamma чи power, швидкість навчання (крок навчання), попередньо навчена модель. Комбінації параметрів в процесі навчання шести моделей зведено у табл. 3.

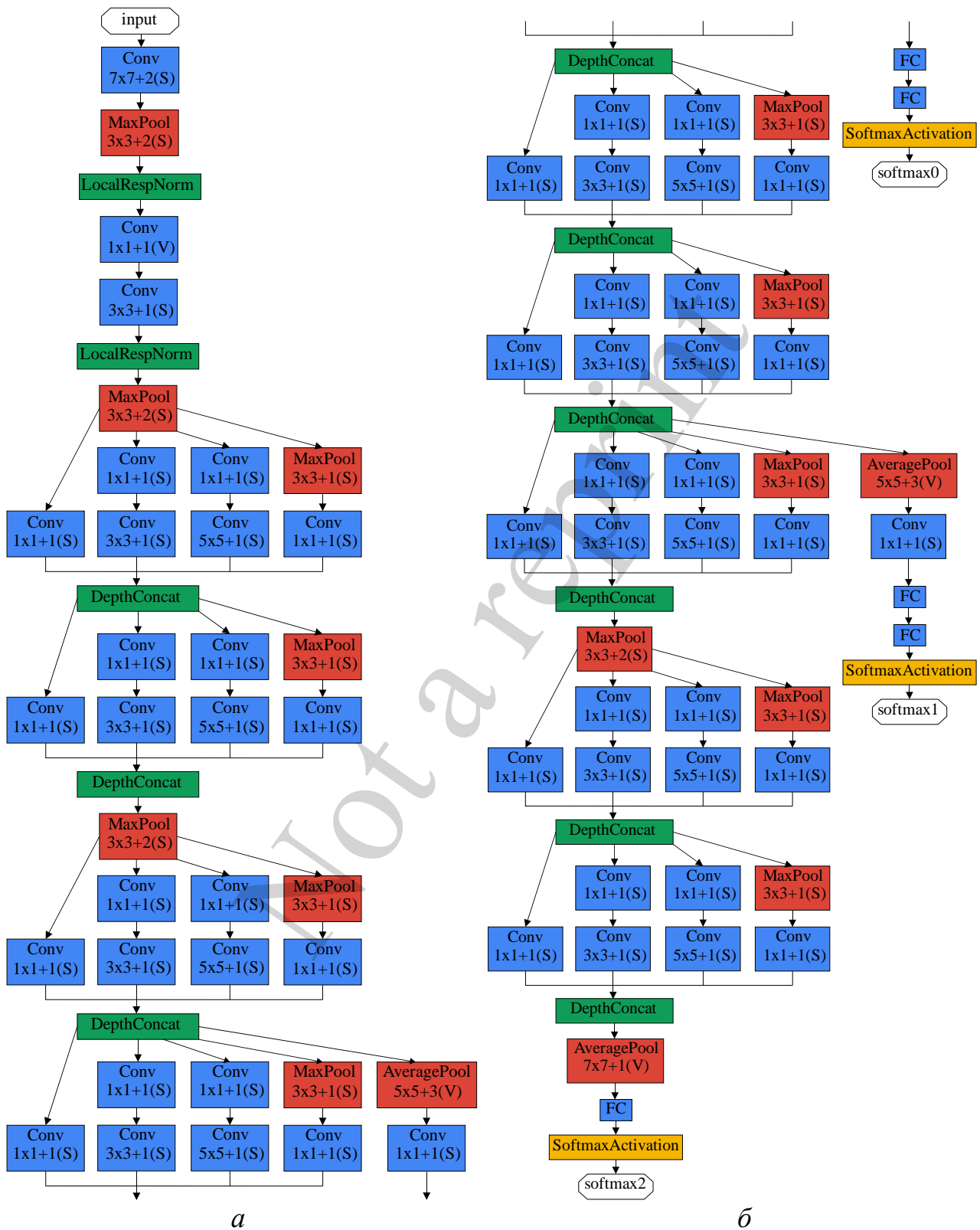


Рис. 1. Архітектура моделі нейронної мережі GoogLeNet: *a* – перша частина; *б* – друга частина

Таблиця 3
Комбінації параметрів в процесі навчання

Навче-на мо-дель	Тривалість навчання (число епох)	Алго-ритм опти-мізації	Вид зміни швидкості навчання, коефіцієнт gamma	Швидкість навчання (крок навчання)	Базується на основі
Inria-1	30	Adam	Експоненційний, 0,99	0,0001	GoogLeNet
Inria-2	30	Adam	Експоненційний, 0,99	0,000075	GoogLeNet
Inria-3	30	Adam	Експоненційний, 0,99	0,00005	GoogLeNet
Inria-4	30	Adam	Експоненційний, 0,99	0,000025	GoogLeNet
Inria-5	100	Adam	Експоненційний, 0,99	0,00001	GoogLeNet
Inria-6	100	Adam	Експоненційний, 0,99	0,000075	GoogLeNet

Алгоритм Adam показує хороші результати оптимізації, зокрема по тривалості навчання, проте не завжди демонструє задовільну збіжність [19]. Тому для навчання моделі з дотриманням балансу між хорошою збіжністю та тривалістю навчання використано різні значення кроку навчання (табл. 3). При хорошій збіжності моделі значення характеристик (1)–(3) мають стабільні значення. В іншому випадку збільшується ризик перенавчання, а значення характеристик (1)–(3) різко змінюються в процесі навчання, що ускладнює практичне використання моделі. Тому моделі з частими та різкими перепадами значень характеристик (1)–(3) не будуть використовуватись у якості попередньо навчених для інших моделей. Для додаткової перевірки на правильність вибраних значень кроку навчання для забезпечення задовільної збіжності та відсутності перенавчання буде виконана перевірка моделі на іншому наборі зображень, який не використовувався для перевіркової вибірки. Значення характеристик (1)–(3) на новому наборі не повинні значно відрізнятись від значень, отриманих для перевіркового набору зображень, що також свідчитиме про адекватність отриманої моделі.

Дослідження проводилось на перевіірочній вибірці, які являють собою набір розмічених зображень INRIA. Особливості: 2 класи об'єктів; знімки у вигляді кольорових зображень з роздільною здатністю 0,3 м загальною площею покриття 810 км², з яких 405 км² для навчання та 405 км² для перевірки.

Значення влучності, чутливості та оцінки усередненої точності на перевіірочній вибірці повинні поступово збільшуватись. Ці параметри, а особливо оцінка усередненої точності, яка включає влучність та чутливість, характеризують адекватність моделі, тобто правильність навчання нейронної мережі та відсутність перенавчання. Валідація результатів навчання може визначатись як поступове збільшення влучності, чутливості та оцінки усередненої точності на перевіірочній вибірці. Кількість епох навчання вибирається з умови отримання найвищої влучності, чутливості та оцінки усередненої точності на перевіірочній вибірці при відсутності суттєвих коливань числових значень. Критерієм доцільності збільшення епох навчання є поступове підвищення влучності, чутливості та оцінки усередненої точності на перевіірочній вибірці. Початок падіння влучності, чутливості та оцінки усередненої точності на перевіірочній вибірці є критерієм перенавчання, відсутність якого є умовою валідації моделі.

5. Результати дослідження розпізнавання об'єктів на зображеннях за допомогою згорткових нейронних мереж

5.1. Дослідження моделей нейромереж для розпізнавання об'єктів на зображеннях з набору INRIA

Дослідження проводилось на попиксельно розмічених зображеннях з набору INRIA. Оскільки розміри зображень в наборі різні, то у разі потреби перевірки моделі на новому зображенні необхідно виконати розмітку наявних об'єктів. Крім того, розрахунок точності відбувається в режимі м'якого реального часу, що вимагає від нейромережі високої швидкодії при обмеженому обсягу пам'яті для забезпечення високої точності. Зображення на вхід нейромережі надходить із 8-бітною колірною палітрою RGB (256 кольорів) розміром не більше 5000×5000 пікселів при роздільній здатності 30 см. Це відповідає поверхні площею до 1500×1500 м. Вихідне зображення формується в форматі TIFF або GeoTIFF. Розміром пакета (batch size) складає 32 з кількістю потоків, рівною 4. Моделі імпортуються у форматі prototxt або protobuf. Час виконання розпізнавання не повинен перевищувати 50 мс для зображення розміру Full HD.

Перевірка ефективності моделей проілюстровано графіками, побудова яких виконувалась автоматично в програмному середовищі DIGITS на основі заданих параметрів, які наведені в табл. 3. Для апаратного прискорення навчання використовувалось середовище Caffe. Кількість значень влучності, чутливості та оцінки усередненої точності рівна кількості епох навчання.

На рис. 2 показано графік перевірки ефективності моделі Inria-1.

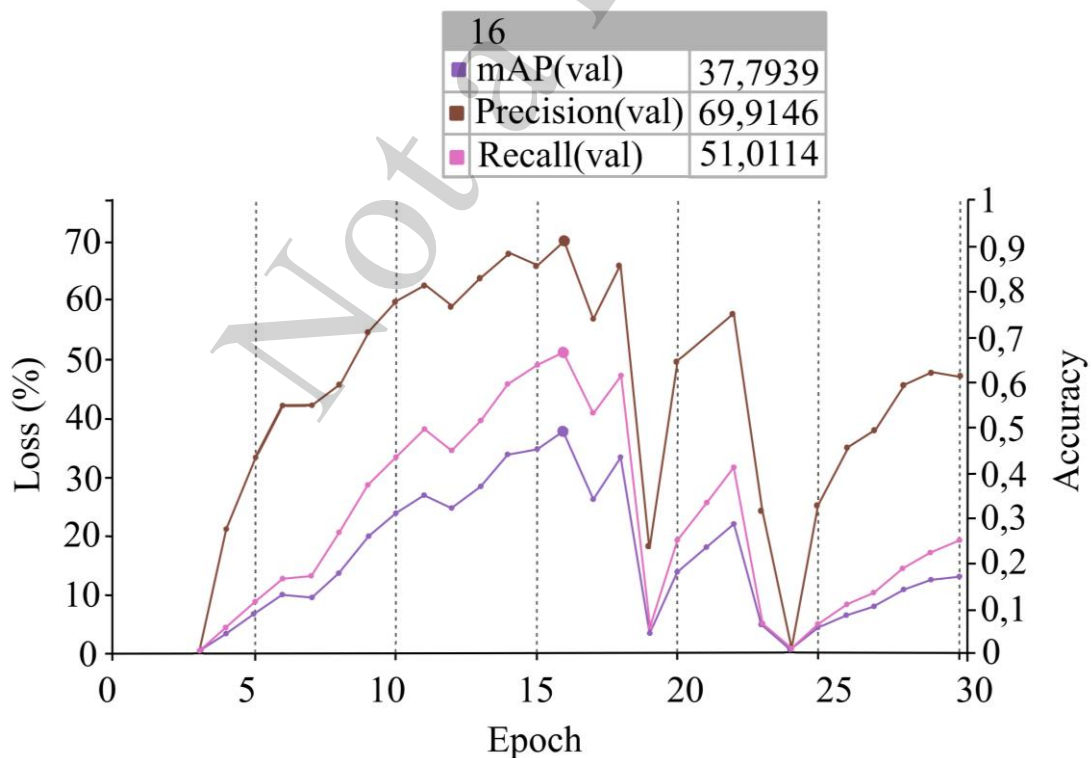


Рис. 2. Графіки зміни на перевіірочній вибірці залежно від епохи для моделі Inria-1: ■ – влучності; ■ – чутливості; ■ – оцінки усередненої точності

З рис. 2 видно, що значення влучності, чутливості та оцінки усередненої точності поступово збільшуються і набувають максимального значення на 16-й епосі навчання. Значення влучності, чутливості та оцінки усередненої точності складають 69,91 %, 51,01 % та 37,79 %, відповідно.

На рис. 3 показано графік перевірки ефективності моделі Inria-2.

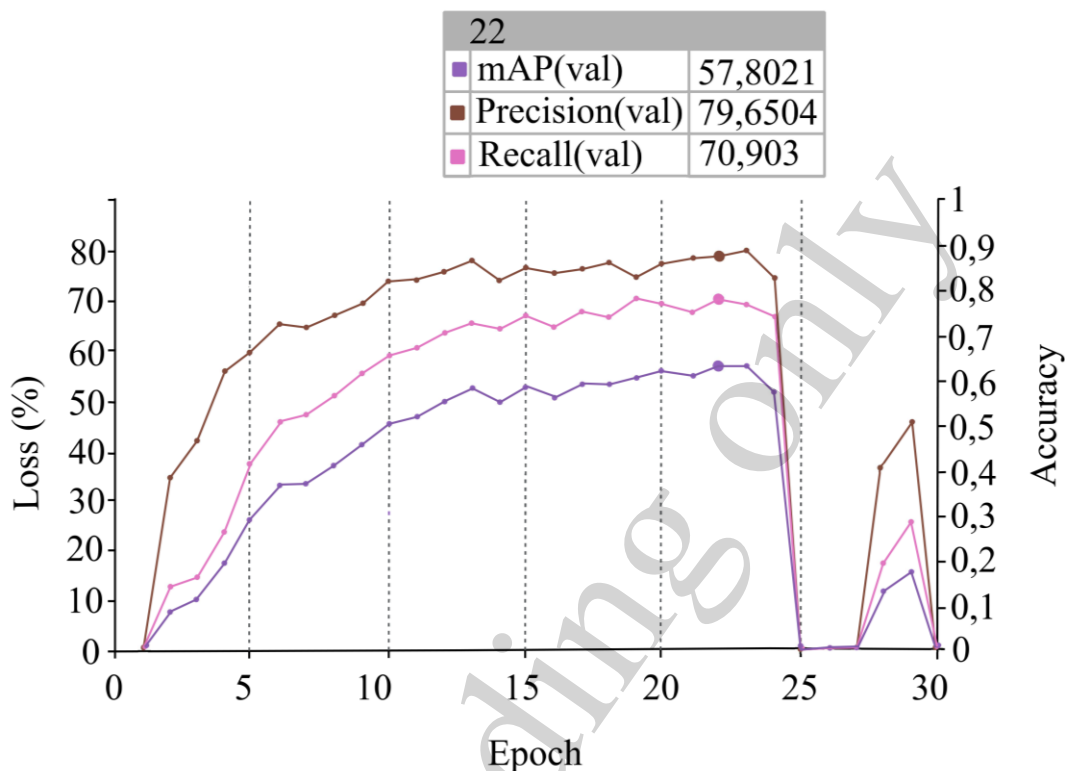


Рис. 3. Графіки зміни на перевіірочній вибірці залежно від епохи для моделі Inria-2: ■ – влучності; ■ – чутливості; ■ – оцінки усередненої точності

З рис. 3 видно, що значення влучності, чутливості та оцінки усередненої точності збільшуються і набувають максимального значення на 23-й епосі для влучності та 22-й для чутливості та оцінки усередненої точності. Значення влучності, чутливості та оцінки усередненої точності для 22-ї епохи складають 79,65 %, 70,90 % та 57,80 %, відповідно.

На рис. 4 показано графік перевірки ефективності моделі Inria-3.

З рис. 4 видно, що значення влучності, чутливості та оцінки усередненої точності поступово збільшуються і набувають максимального значення на 30-й епосі навчання. Значення влучності, чутливості та оцінки усередненої точності складають 79,38 %, 68,08 % та 55,41 %, відповідно.

На рис. 5 показано графік перевірки ефективності моделі Inria-4.

З рис. 5 видно, що значення влучності, чутливості та оцінки усередненої точності поступово збільшуються і набувають максимального значення на 29-й епосі навчання. Значення влучності, чутливості та оцінки усередненої точності складають 77,44 %, 66,54 % та 52,59 %, відповідно.

На рис. 6 показано графік перевірки ефективності моделі Inria-5.

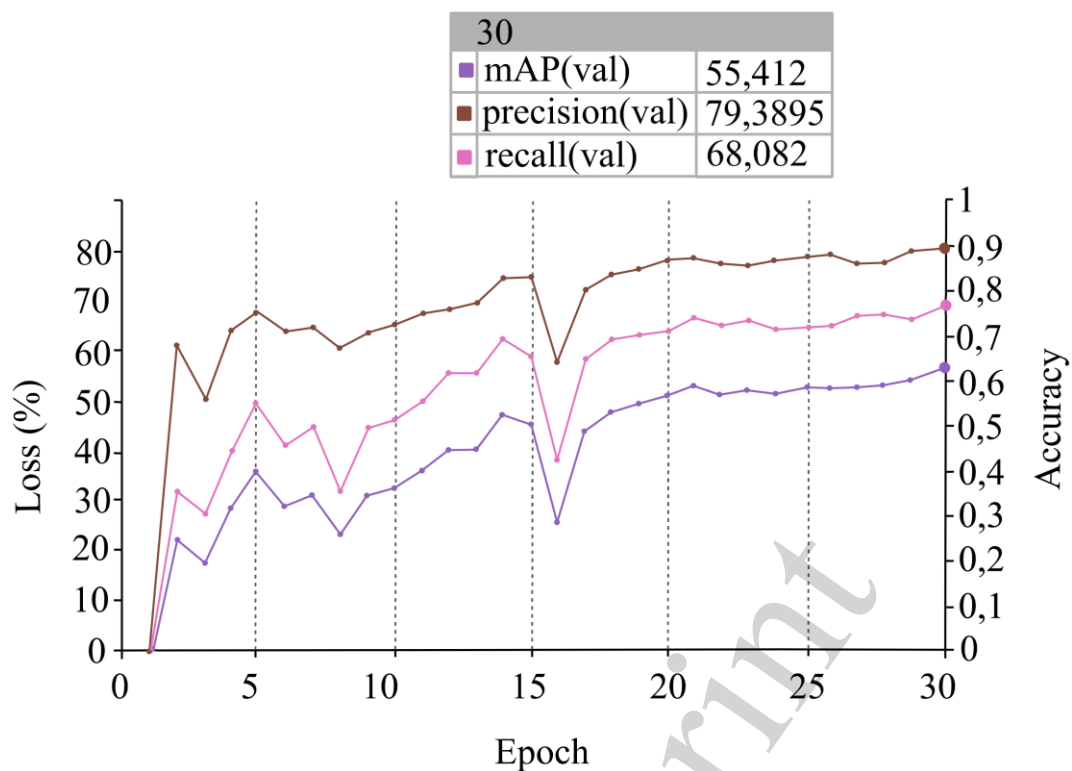


Рис. 4. Графіки зміни на перевіірочній вибірці залежно від епохи для моделі Inria-3: ■ – влучності; ■ – чутливості; ■ – оцінки усередненої точності

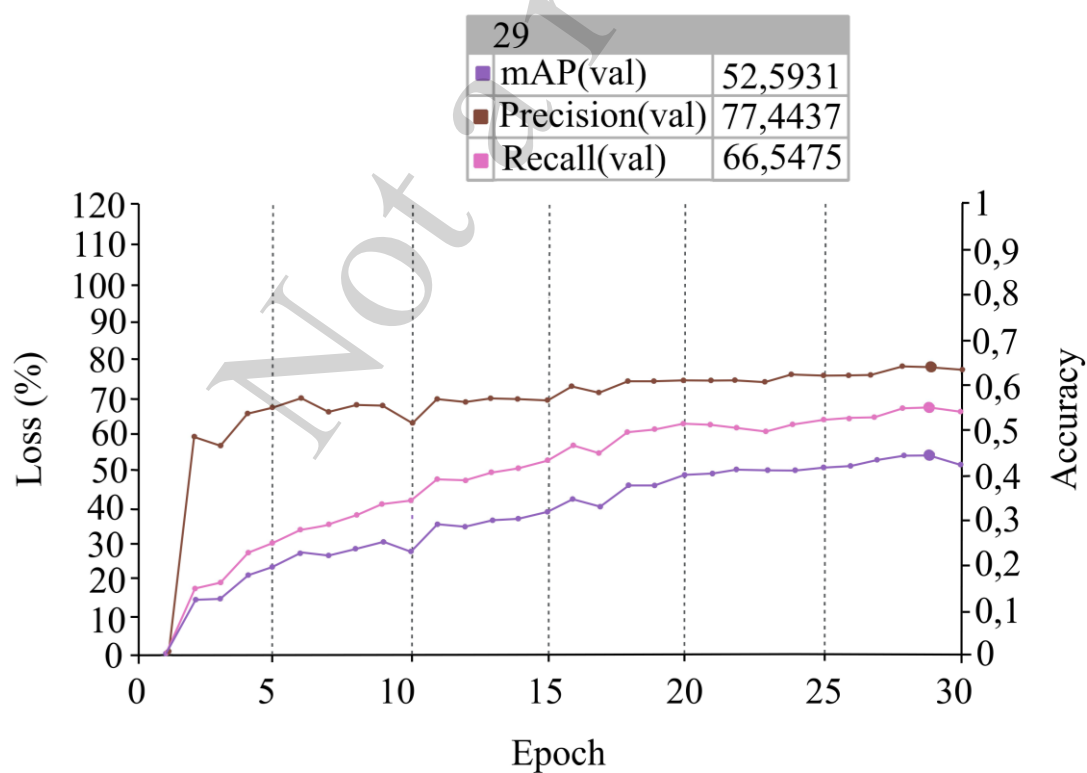


Рис. 5. Графіки зміни на перевіірочній вибірці залежно від епохи для моделі Inria-4: ■ – влучності; ■ – чутливості; ■ – оцінки усередненої точності

З рис. 6 видно, що значення влучності, чутливості та оцінки усередненої точності поступово збільшуються і набувають максимального значення на 94-й епосі навчання. Значення влучності, чутливості та оцінки усередненої точності складають 77,55 %, 63,76 % та 50,83 %, відповідно.

На рис. 7 показано графік перевірки ефективності моделі Inria-6.

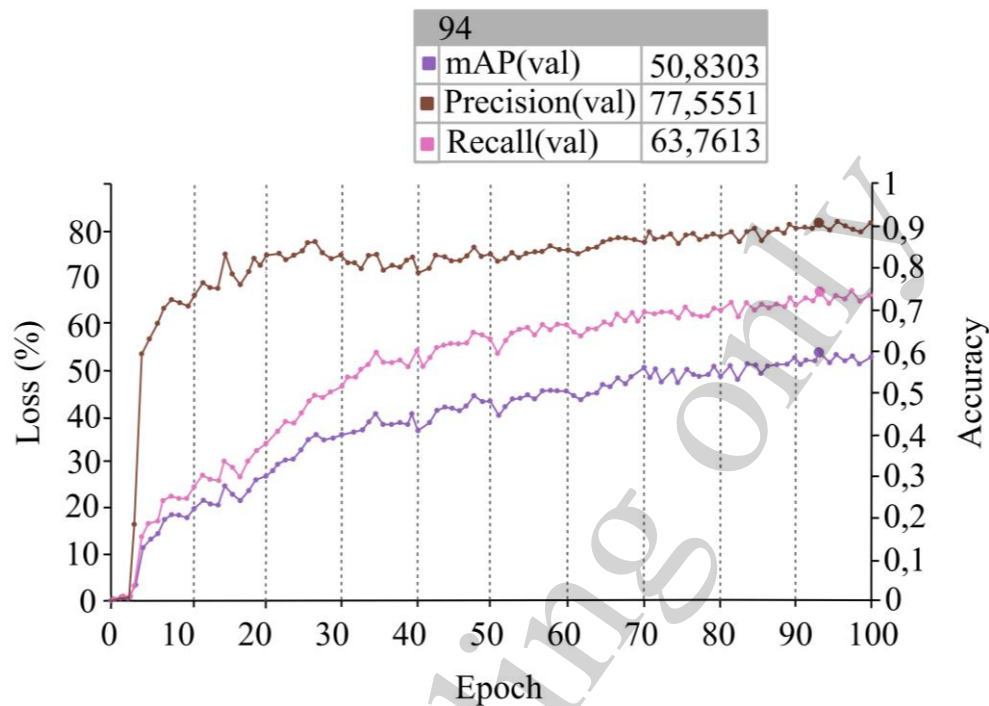


Рис. 6. Графіки зміни на перевіірочній вибірці залежно від епохи для моделі Inria-5: ■ – влучності; ■ – чутливості; ■ – оцінки усередненої точності

З рис. 7 видно, що значення влучності, чутливості та оцінки усередненої точності збільшуються і набувають максимального значення на 23-й епосі для влучності та 22-й для чутливості та оцінки усередненої точності. Значення влучності, чутливості та оцінки усередненої точності для 22-ї епохи складають 79,65 %, 70,90 % та 57,80 %, відповідно.

В результаті досліджень отримані результати перевірки ефективності шести моделей занесено в табл. 4.

Таблиця 4

Результати досліджень ефективності моделей із різними параметрами

Навчена модель	Епоха з найкращим результатом/кількість епох	Оцінка усередненої точності, %	Влучність, %	Чутливість, %
Inria-1	16/30	37,79	69,91	51,01
Inria-2	22/30	57,80	79,65	70,90
Inria-3	30/30	55,41	79,38	68,08
Inria-4	29/30	52,59	77,44	66,54
Inria-5	94/100	50,83	77,55	63,76
Inria-6	22/100	57,80	79,65	70,90

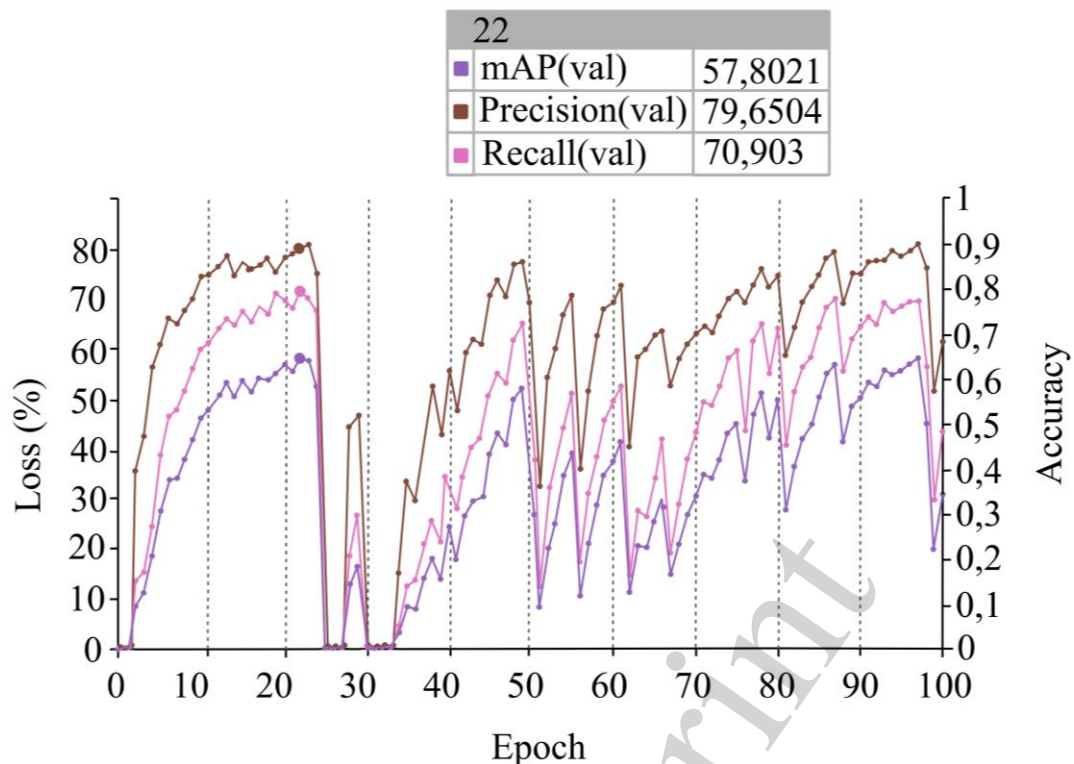


Рис. 7. Графіки зміни на перевіірчній вибірці залежно від епохи для моделі Inria-6: ■ – влучності; ■ – чутливості; ■ – оцінки усередненої точності

З табл. 4 видно, що найбільшу оцінку усередненої точності при відсутності різких стрибків у показниках демонструє модель Inria-3 з 55,41 % при 30 епохах навчання та швидкості навчання 0,00005.

Таким чином, Inria-3 використано як основу для навчання нової моделі Inria-7 при 30 епохах з експоненційною зміною швидкості навчання, яка складає 0,000025, коефіцієнтом $\gamma = 0,99$ та видом оптимізації Adam.

На рис. 8 показано графік перевірки ефективності моделі Inria-7.

З рис. 8 видно, що значення влучності, чутливості та оцінки усередненої точності поступово збільшуються і набувають максимального значення на 23-й епосі навчання. Значення влучності, чутливості та оцінки усередненої точності складають 82,12 %, 72,69 % та 60,77 %, відповідно.

Таким чином, значення оцінки усередненої точності зросло із 55,41 % до 60,77 %.

Дана модель показує хороші показники зростання оцінки усередненої точності та стабільність результатів, тому її використано для навчання Inria-8 та Inria-9 (поліноміальна зміна швидкості навчання) при тривалості навчання збільшеній до 100 (табл. 5).

На рис. 9 показано графік перевірки ефективності моделі Inria-8.

З рис. 9 видно, що значення влучності, чутливості та оцінки усередненої точності поступово збільшуються і набувають максимального значення на 97-й епосі навчання. Значення влучності, чутливості та оцінки усередненої точності складають 84,15 %, 74,00 % та 63,22 %, відповідно.

На рис. 10 показано графік перевірки ефективності моделі Inria-9.

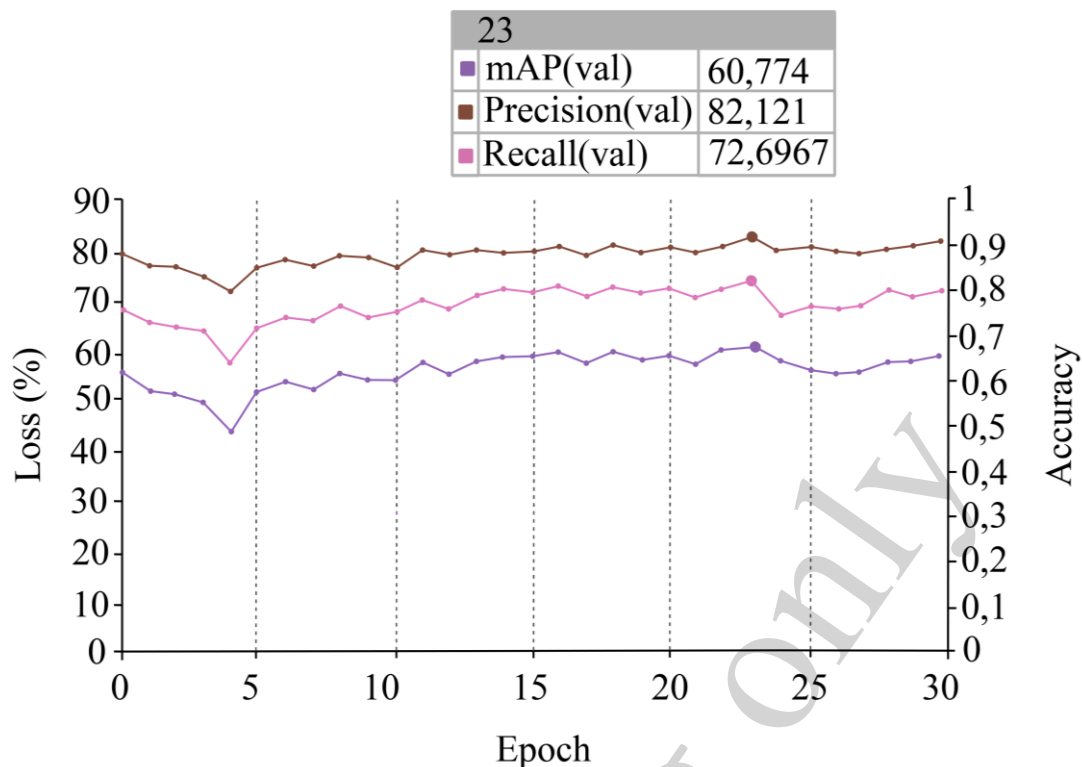


Рис. 8. Графіки зміни на перевіірочній вибірці залежно від епохи для моделі Inria-7: ■ – влучності; ■ – чутливості; ■ – оцінки усередненої точності

Таблиця 5
Параметри, які змінювались в процесі навчання

Навчена модель	Тривалість навчання (число епох)	Алгоритм оптимізації	Вид зміни швидкості навчання, коефіцієнт gamma/power (для поліноміального)	Швидкість навчання	Базується на основі
Inria-8	100	Adam	Експоненційний, 0,99	0,00001	Inria-7
Inria-9	100	Adam	Поліноміальний, 3	0,00005	Inria-7

З рис. 10 видно, що значення влучності, чутливості та оцінки усередненої точності збільшуються і набувають максимального значення на 24-й епосі для чутливості та 45-й для влучності та оцінки усередненої точності. Значення влучності, чутливості та оцінки усередненої точності для 45-ї епохи складають 85,68 %, 75,59 % та 65,70 %, відповідно.

В результаті досліджень отримані результати перевірки ефективності трьох моделей занесено в табл. 6.

З табл. 6 видно, що найбільшу оцінку усередненої точності при відсутності різких стрибків у показниках демонструє модель Inria-9. Це спостерігається при 100 епохах з поліноміальною зміною швидкості навчання, яка складає 0,00005, коефіцієнтом power – 3 та видом оптимізації Adam.

Для цієї моделі вдалось збільшити оцінку усередненої точності з 60,77 % до 65,70 %.

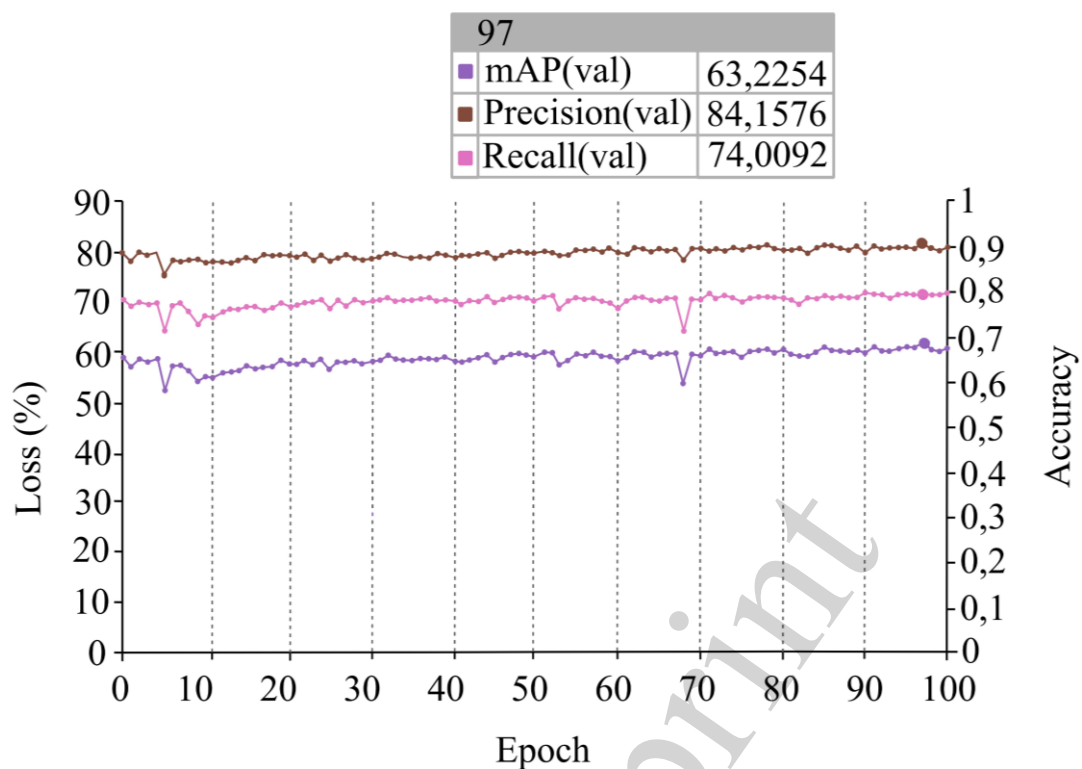


Рис. 9. Графіки зміни на перевіірочній вибірці залежно від епохи для моделі Inria-8: ■ – влучності; ■ – чутливості; ■ – оцінки усередненої точності

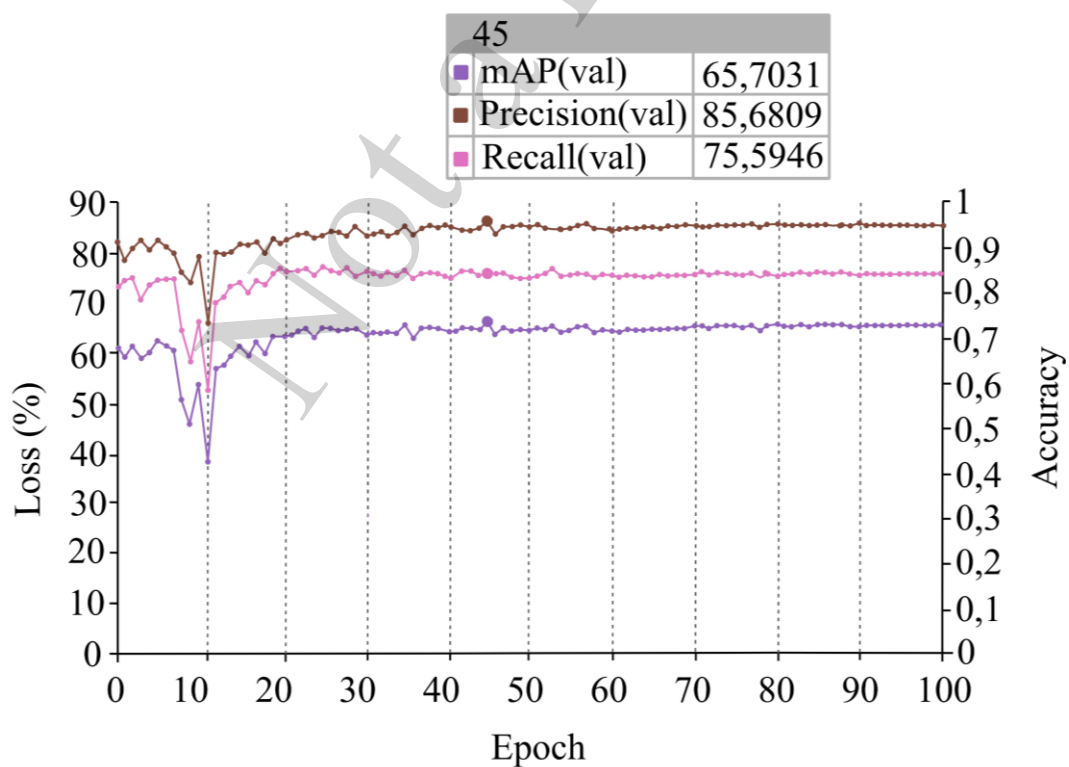


Рис. 10. Графіки зміни на перевіірочній вибірці залежно від епохи для моделі Inria-9: ■ – влучності; ■ – чутливості; ■ – оцінки усередненої точності

Таблиця 6

Результати досліджень ефективності моделей із різними параметрами

Навчена модель	Епоха з найкращим результатом/кількість епох	Оцінка усередненої точності, %	Влучність, %	Чутливість, %
Inria-7	23/30	60,77	82,12	72,69
Inria-8	97/100	63,22	84,15	74,00
Inria-9	45/100	65,70	85,68	75,59

5. 2. Оцінка навченої моделі Inria-9 для розпізнавання об'єктів на зображеннях

На практиці розпізнавання об'єктів на зображеннях є складовою моніторингу навколишнього середовища за допомогою БПЛА, що вимагає високої точності для керування та орієнтації в просторі. Тому модель з найбільшою оцінкою усередненої точності Inria-9 потрібно донавчати використовуючи значення параметрів, визначені як оптимальні на основі дослідження (табл. 3):

- швидкість навчання – 0,000025;
- тривалість навчання (число епох) – 100;
- алгоритм оптимізації – Adam;
- вид зміни швидкості навчання – поліноміальний;
- коефіцієнт power – 0,25;
- попередньо навчена модель – Inria-7.

Таким чином, отримано модель Inria-10, результати перевірки якої показано на рис. 11.

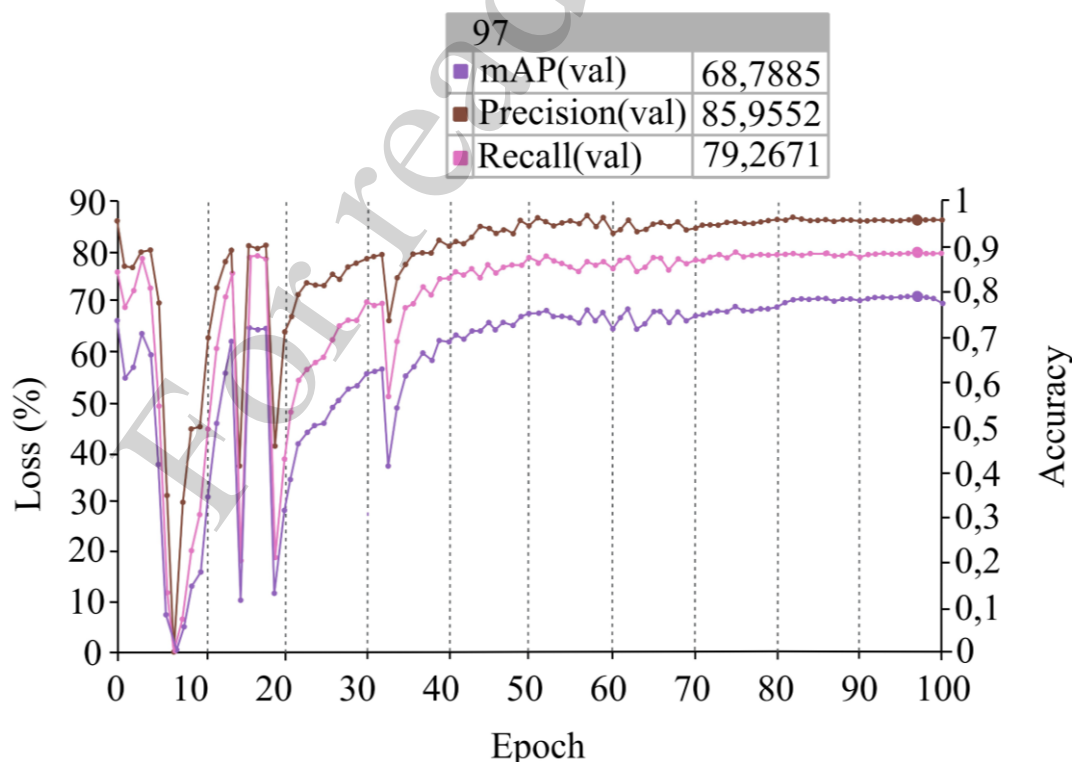


Рис. 11. Графіки зміни на перевіірочній вибірці залежно від епохи для моделі Inria-10: ■ – влучності; ■ – чутливості; ■ – оцінки усередненої точності

З рис. 11 видно, що значення влучності, чутливості та оцінки усередненої точності поступово збільшуються і набувають максимального значення на 97-й епосі навчання. Значення влучності, чутливості та оцінки усередненої точності складають 85,95 %, 79,26 % та 68,78 %, відповідно.

В результаті досліджень отримані результати перевірки ефективності всіх десяти моделей занесено в табл. 7.

Таблиця 7

Результати досліджень ефективності моделей із різними параметрами

Навчена модель	Епоха з найкращим результатом/кількість епох	Вид зміни швидкості навчання	Швидкість навчання	Базується на основі	Оцінка усередненої точності, %
Inria-1	16/30	Експоненційний	0,0001	GoogLeNet	37,79
Inria-2	22/30	Експоненційний	0,000075	GoogLeNet	57,80
Inria-3	30/30	Експоненційний	0,00005	GoogLeNet	55,41
Inria-4	29/30	Експоненційний	0,00025	GoogLeNet	52,59
Inria-5	94/100	Експоненційний	0,00001	GoogLeNet	50,83
Inria-6	22/100	Експоненційний	0,000075	GoogLeNet	57,80
Inria-7	23/30	Експоненційний	0,000025	Inria-3	60,77
Inria-8	97/100	Експоненційний	0,00001	Inria-7	63,22
Inria-9	45/100	Поліноміальний	0,00005	Inria-7	65,70
Inria-10	97/100	Поліноміальний	0,000025	Inria-9	68,78

З табл. 7 видно, що серед десяти навчених моделей найвищою оцінкою усередненої точності володіє Inria-10. Для цієї моделі вдалось збільшити оцінку усередненої точності з 55,41 % (Inria-3) до 60,77 % (Inria-7), потім до 65,70 % (Inria-9) і, нарешті, до 68,78 %. Подальшого збільшення можна досягти шляхом експериментів зі зміни архітектури нейромережі, більш старанного вибору зображень з набору та комбінацією навчань на різних наборах даних, але це вимагає значних обчислювальних ресурсів.

На рис. 12 показано розпізнавання будівель на зображеннях з відеокамери БПЛА для моделі Inria-10 в програмному середовищі DIGITS.

З прикладу (рис. 12) можна зробити висновок, що мережа розпізнає майже всі будівлі. Нерозпізнаними залишились такі споруди як сараї, теплиці, недобудови, а також будівлі, які були присутні на фотографії частково, перекривалися деревами або, в зв'язку з близьким розташуванням, розпізнавалися як одна будівля. Кількість нерозпізнаних будівель підтверджує експериментальну точність близько 70 %. При цьому помилкового віднесення об'єктів, які не є будівлями, до класу «будівля» не зафіксовано.

Таким чином, якість роботи моделі Inria-10 суттєво залежить від візуальних розмірів потрібного об'єкту, освітлення, кута зйомки, наявності предметів, які заважають огляду. Проте при близькій зйомці за гарного освітлення будівлі майже гарантовано розпізнаються. Тому дану модель можна використовувати

для контролю сільськогосподарських господарств, створення ортофотопланів, складання карт полів, моніторингу територій, вирішення завдань, пов'язаних з кадастром та землеустроєм, тощо.



Рис. 12. Приклад розпізнавання будівель для моделі Inria-10

Наявні значення швидкодії дозволяють порівняти модель Inria-10 з іншими [2–10], проте є сенс проводити порівняння з близькими по архітектурі моделями, які отримані при навчанні на аналогічній базі. Тому порівняння проведено по критерію оцінки усередненої точності з деякими відомими моделями на основі GoogLeNet, розробленими за схожими параметрами, навченими на основі зображень, отриманих з камер БПЛА. Такими моделями було обрано GoogLeNet-like (Швейцарія), InceptionResNetV2 (Туреччина), U-Net

InceptionResNetV2 (Туреччина) [20, 21]. Результати оцінки усередненої точності моделей занесено в табл. 8.

Таблиця 8
Результати оцінки точності моделей

Назва моделі	Оцінка усередненої точності, %
GoogLeNet-like	70
InceptionResNetV2	72
U-Net InceptionResNetV2	73
Inria-10	75

З табл. 8 видно, що найбільшу оцінку усередненої точності, а саме 75 %, демонструє модель Inria-10, оскільки інші навчалися на основі зображень, отриманих з камер БПЛА, які не входили в набір INRIA.

Виконано оцінку адекватності, достовірності та збіжності моделі Inria-10 у порівнянні з іншими [20, 21]. Для цього проведено розпізнавання 100 розмічених вручну зображень із набору NVidia Aerial Drone Dataset (США) [22]. Даний набір обраний, оскільки зображення в цьому наборі отримані за інших умов зйомки, ніж в наборі INRIA. Розрахунки виконувались в середовищі Jupiter Notebook на мові Python. Усереднені результати зведено до табл. 9.

Таблиця 9
Результати перевірки для оцінки достовірності, адекватності та збіжності моделі

Назва моделі	Влучність (precision), %	Чутливість (recall), %	Оцінка усередненої точності (mAP), %
GoogLeNet-like	81	76	63
InceptionResNetV2	82	79	64
U-Net InceptionResNetV2	84	79	65
Inria-10	87	81	67

З табл. 9 видно, що найбільшу оцінку усередненої точності, а саме 67 %, демонструє розроблена модель Inria-10. Це свідчить про високу достовірність моделі Inria-10.

Об'єкти на зображеннях для перевірки розмічені на певні класи, згідно яких відбувається розпізнавання моделлю заданих об'єктів (будівель) на зображенні. При експериментальній перевірці моделей були хибно позитивні (наявність певного класу на зображенні за його відсутності) та хибно негативні (відсутність певного класу на зображенні за його присутності) результати при позначенні класів. Частка зображень із хибно розміченими класами складає 15 %, а частка зображень із правильно розміченими класами – 85 %. Дані результати дозволяють за формулами (1)–(3) обчислити влучність, чутливість та оцінку усередненої точності. Значення влучності складає 87 %, що свідчить про збіжність моделі. Отримане значення чутливості, яке складає 81 % свідчить про до-

стовірність моделі. Значення оцінки усередненої точності складає 67 %, що свідчить про адекватність моделі.

6. Обговорення результатів дослідження розпізнавання об'єктів на зображеннях за допомогою згорткових нейронних мереж

Результати досліджень показують, що навчена модель Inria-10 демонструє високу точність розпізнавання об'єктів на зображеннях (рис. 11). Це пояснюється вибором оптимальних параметрів нейронної мережі, а також внесенням згорткового шару в стандартну архітектуру нейромережі. Inria-10 базується на Inria-9. Ця модель продемонструвала найкращі значення оцінки усередненої точності (табл. 6) в процесі навчання на основі набору INRIA (рис. 10). Це пояснюється вибором поліноміальної зміни швидкості навчання (табл. 5). Тому саме Inria-9 було обрано для донавчання з оптимальними параметрами нейронної мережі. Створену таким чином модель Inria-10 можна використовувати для розпізнавання об'єктів на реальних зображеннях (рис. 12), висока точність якої визначає ефективність системи керування БПЛА. Результати порівняння оцінки усередненої точності розпізнавання об'єктів на зображеннях для Inria-10 та інших подібних моделей занесено в табл. 8. Inria-10 в порівнянні з іншими демонструє високі значення оцінки усередненої точності розпізнавання об'єктів на зображеннях, що вказує на адекватність цієї моделі та відсутність перенавчання.

Точність та швидкодія розробленої моделі нейронної мережі Inria-10 мають більші значення, ніж аналогічні, отримані в [20], на 2–4 % та 20–50 %, відповідно. При цьому процес розпізнавання не вимагає значних обчислювальних ресурсів на етапі використання моделі. Порівняно з [21] дана модель має вищу точність розпізнавання об'єктів на зображеннях на 3 %. Це досягнуто за рахунок додавання шарів даних, перетворення, нормалізації даних, похибки, обчислення усередненої похибки та параметрів для доповнення даних, а також FCN і видалення шарів вхідних/вихідних даних та pooling шару. Достовірність, адекватність та збіжність розробленої моделі нейронної мережі Inria-10 є порівнювана (табл. 9) із іншими моделями [20, 21] та не поступається їм. Значення влучності більше на 3–6 %, що свідчить про збіжність моделі. Значення чутливості більше на 2–5 %, що свідчить про достовірність моделі. Значення оцінки усередненої точності більше на 2–4 %, що свідчить про адекватність моделі.

Оскільки навчання моделі нейромережі проводилось для зображень з набору INRIA, то високі значення точності розпізнавання притаманні зображенням, отриманим із відеокамери дрона, зазвичай через високу контрастність груп пікселів. Для інших видів зображень, імовірно, точність не буде настільки високою. Це потребує додаткових досліджень.

До недоліків варто віднести затратність часу та обчислювальних ресурсів на етапі навчання нейромережі. Даний недолік можна подолати використанням паралельних графічних обчислень з допомогою технології CUDA та використанням як попередньо навченої нейромережі більш компактно, наприклад, MobileNet.

Розвиток дослідження даної моделі може полягати у подальшому підвищенні точності, швидкодії та зменшенні обсягів обчислювальних ресурсів. Це

вимагатиме складного математичного моделювання, врахування предметної області застосування та розробки програмних модулів для конкретної системи.

7. Висновки

1. Досліджено моделі Inria-1, Inria-2, Inria-3, Inria-4, Inria-5, Inria-6, Inria-7, Inria-8, Inria-9 нейромереж на основі набору INRIA. Встановлено, що найбільшу оцінку усередненої точності демонструє модель Inria-9 з 68,78 % при швидкості навчання 0,000025 на основі Adam при поліноміальній зміні швидкості навчання з коефіцієнтом power – 0,25. Найменше значення оцінки усередненої точності складає 37,79 % для моделі Inria-1, яка використовує експоненційну зміну швидкості навчання. Це означає, що більша точність навчання забезпечується за рахунок поліноміальної зміни швидкості навчання.

2. Отримано значення оцінки усередненої точності моделі Inria-10, створеної на основі попередньо навченої моделі Inria-9 для розпізнавання об'єктів на зображеннях з набору INRIA з визначеними на основі дослідження параметрами. Це значення є досить високим, адже поступово збільшуються і набуває максимального значення на 97-й епосі навчання. Значення влучності, чутливості та оцінки усередненої точності складають 85,95 %, 79,26 % та 68,78 %, відповідно. Отримані значення дозволяють стверджувати про правильність вибору архітектури мережі та підбору параметрів. Це дає можливість використовувати дану модель для практичних задач розпізнавання об'єктів на зображеннях, наприклад, в автопілотах, в системах уникнення зіткнень з іншими БПЛА, для машинного зору, аналізу сільськогосподарської інфраструктури тощо.

Література

1. Білинський, Й. Й., Книш, Б. П., Кулик, Я. А. (2017). Методика оцінювання якості роботи фільтрів приглушення шумів в пакеті MathCAD. Вісник Хмельницького національного університету, 3, 125–130. URL: <http://ir.lib.vntu.edu.ua/bitstream/handle/123456789/23238/47857.pdf?sequence=2&isAllowed=y>
2. Gall, J., Razavi, N., Van Gool, L. (2012). An Introduction to Random Forests for Multi-class Object Detection. Outdoor and Large-Scale Real-World Scene Analysis, 243–263. doi: https://doi.org/10.1007/978-3-642-34091-8_11
3. Viola, P., Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001. doi: <https://doi.org/10.1109/cvpr.2001.990517>
4. Weiming Hu, Wei Hu, Maybank, S. (2008). AdaBoost-Based Algorithm for Network Intrusion Detection. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 38 (2), 577–583. doi: <https://doi.org/10.1109/tsmcb.2007.914695>
5. Shang, W., Sohn, K., Almeida, D., Honglak, L. (2016). Understanding and Improving Convolutional Neural Networks via Concatenated Rectified Linear Units. Proceedings of The 33rd International Conference on Machine Learning, 48, 2217–2225. URL: <http://proceedings.mlr.press/v48/shang16.html>

6. Simonyan, K., Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. ICLR. URL: <https://arxiv.org/pdf/1409.1556.pdf>
7. Redmon, J., Divvala, S., Girshick, R., Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi: <https://doi.org/10.1109/cvpr.2016.91>
8. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D. et. al. (2015). Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi: <https://doi.org/10.1109/cvpr.2015.7298594>
9. Prathap, G., Afanasyev, I. (2018). Deep Learning Approach for Building Detection in Satellite Multispectral Imagery. 2018 International Conference on Intelligent Systems (IS). doi: <https://doi.org/10.1109/is.2018.8710471>
10. Wu, K., Chen, Z., Li, W. (2018). A Novel Intrusion Detection Model for a Massive Network Using Convolutional Neural Networks. IEEE Access, 6, 50850–50859. doi: <https://doi.org/10.1109/access.2018.2868993>
11. Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P. (2017). Can semantic labeling methods generalize to any city? The inria aerial image labeling benchmark. 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). doi: <https://doi.org/10.1109/igarss.2017.8127684>
12. Knysh, B., Kulyk, Y. (2021). Development of an image segmentation model based on a convolutional neural network. Eastern-European Journal of Enterprise Technologies, 2 (2 (110)), 6–15. doi: <https://doi.org/10.15587/1729-4061.2021.228644>
13. Krizhevsky, A., Sutskever, I., Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. NIPS'12: Proceedings of the 25th International Conference on Neural Information Processing Systems, 1097–1105. URL: <https://papers.nips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>
14. Zeiler, M. D., Fergus, R. (2014). Visualizing and Understanding Convolutional Networks. Lecture Notes in Computer Science, 818–833. doi: https://doi.org/10.1007/978-3-319-10590-1_53
15. Deep Learning: GoogLeNet Explained. Towards Data Science. URL: <https://towardsdatascience.com/deep-learning-googlenet-explained-de8861c82765>
16. Tao, A., Barker, J., Sarathy, S. (2016). DetectNet: Deep Neural Network for Object Detection in DIGITS. NVidia developer blog. URL: <https://developer.nvidia.com/blog/detectnet-deep-neural-network-object-detection-digits>
17. Kingma, D. P., Ba, J. (2015). Adam: a method for stochastic optimization. ICLR 2015. URL: <https://arxiv.org/pdf/1412.6980.pdf>
18. Кветний, Р. Н., Маслій, Р. В., Кириленко, О. М. (2020). Виявлення та класифікація об'єктів дорожнього руху при використанні середовища DIGITS. Вісник Хмельницького національного університету, 1 (39), 14–20. doi: <https://doi.org/10.31649/1681-7893-2020-39-1-14-20>
19. Wilson, A. C., Roelofs, R., Stern, M., Srebro, N., Recht, B. (2017). The marginal value of adaptive gradient methods in machine learning. 31st Conference on Neural Information Processing Systems (NIPS 2017). URL: <https://arxiv.org/pdf/1705.08292v2.pdf>

20. Guo, Z., Chen, Q., Wu, G., Xu, Y., Shibasaki, R., Shao, X. (2017). Village Building Identification Based on Ensemble Convolutional Neural Networks. *Sensors*, 17 (11), 2487. doi: <https://doi.org/10.3390/s17112487>
21. Erdem, F., Avdan, U. (2020). Comparison of Different U-Net Models for Building Extraction from High-Resolution Aerial Imagery. *International Journal of Environment and Geoinformatics*, 7 (3), 221–227. doi: <https://doi.org/10.30897/ijgeo.684951>
22. Nvidia Aerial Drone Dataset. URL: <https://nvidia.box.com/shared/static/ft9cc5yjvrbhkh07wcivu5ji9zola6i1.gz>

Not a reprint